Speedup of Learning for Active Learning Method by using Supervised Data

Yoshitaka SAKURAI Nakaji HONDA

Department of systems Engineering, University of Electro-Communications (UEC)
1-5-1, Chofugaoka, Chofu-si, Tokyo, 182-8585 Japan
E-mail:ysakurai@fs.se.uec.ac.jp

Abstract— We propose the Active Learning Method(ALM) for acquiring control knowledge actively by trial and error, and it proves about the usefulness. But, Learning by trial and error takes much time for Learning. If it is the case where supervised data can be obtained, control knowledge can be acquired having used it more quickly. Then, supervised learning is performed by ALM. Even when it is difficult to obtain the optimal supervised data, by learning by trial and error based on imperfect supervised data, Improvement in learning speed is aimed at. In this paper, the learning simulation of an action policy for the control problem of double pendulum is performed, and when imperfect supervised data is given compared with the case where there is no knowledge, it verifies how much of speedup is obtained.

I. Introduction

We have proposed the Active Learning Method (ALM) for acquiring control knowledge actively by trial and error [2]. This is a unsupervised learning method. In this paper, The speed rise of ALM learning is aimed at using supervised data.

About machine learning, the approach from various directions, such as an information theory, statistics, statistics physics, and soft computing, is tried. And the most of these need the supervised data to learn. However, an example with it difficult to acquire a supervised data. So a unsupervised learning method is also searched for. Moreover, in unsupervised learning method, the solution superior to the solution which man's expert got may be discovered by learning by trial and error. Then, we proposed ALM for acquiring control knowledge actively by trial and error, and it has proved about the usefulness [1][2].

However, the learning by trial and error takes much time for search a solution. If it is the case where supervised data can be obtained, control knowledge can be acquired having used it more quickly. So, In order to shorten learning time, supervised learning is performed in ALM. Even when it is difficult to obtain the optimal supervised data, improvement in the learning speed of ALM is aimed at by performing learning by trial and error based on imperfect supervised data.

In this paper, the learning simulation of an action policy in the control problem of double pendulum is performed, and when imperfect supervised data is given compared with the case where there is no knowledge, it verifies whether improvement in the speed of how much is obtained.

II. Active Learning Method (ALM)

A. Active Learning Method

In learning of human, one of the most important elements is "experience." A judgment of human is made in many cases based on the past experience. It is thought that man's movement control is one of this, and the rule like a pattern is learned experientially. Then, as a humanlike learning method, we have proposed the Active Learning Method (ALM). ALM is modeling a system in pattern information based on data. And, using fuzzy concept ALM model a system.

When human learn about a system, at first information is collected. And, it finds out a tendency from much information which looks disorderly.

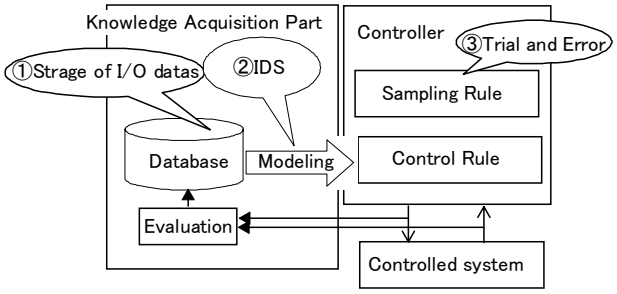


Fig. 1 System of ALM

It is the same also at ALM, and when learning the action of a certain system, it starts with grasping the input-and-output information on the system. 1--At first the input-and-output data of the target system is collected. (in the study from zero, operation is tried at random. when there is supervised data, It is used.) And, these data are evaluated and saved. 2--Model system by these data. 3--Save the model.

And, collect Input-and-output data by trial and error. 4--Model system by previous model and new data. 5--repeat these steps.

B. Modeling by the fuzzy process

ALM grasps the feature of a system by Narrow path – continues path on a data plane. Narrow path is extracted from input-and-output data by project data on the data plane. And, in ALM, the system of a multi input multi output (MIMO) is divided into the system of a single input single output (SISO), and it is expressing by combining it (Fig.2). Ink Drop Spread (IDS) is used for extraction of Narrow path of this SISO system.

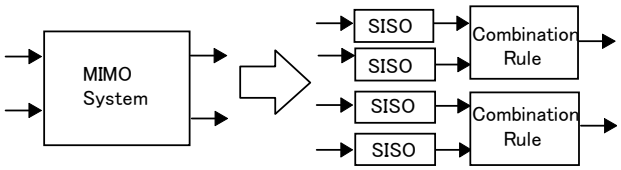


Fig. 2 Division and integration of a system

The basic concept of IDS is to extract the system properties (narrow path) from the input-output data by using fuzzy process. Here, all data on the data plane are supposed to be light sources (Fig.3). When irradiated from the top perpendicularly above the data, the light interferes with each other and the illuminated pattern appears to show light and darkness. That is, the part where many lights fall is lighter than other part. By combining the light parts continuously, a kind of narrow path expressing the input-output relations can be obtained.

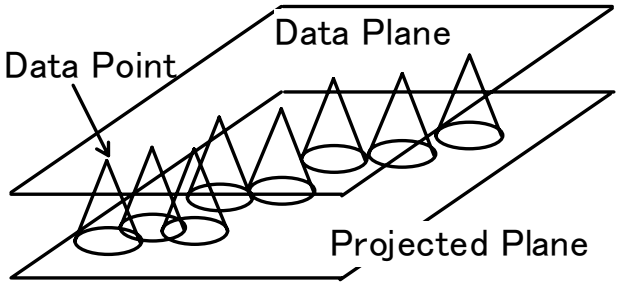


Fig.3 IDS (Ink Drop Spread)

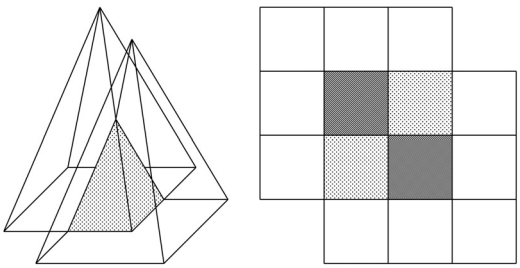


Fig.4 Irradiation Pyramid

Here, the irradiation pyramid like Fig.4 can be handled as the margin of allowance of vagueness of the data, that is, three-dimensional membership function. From this, this method is said to include vagueness at the time of input. Further by applying IDS, information about spread of the irradiation pattern together with the narrow path can be obtained. The information of this spread enables detection of uncertain part of the input-output relation and expression of the probability action.

For example, modeling of two inputs and one output system $y = f(x_1, x_2)$ is shown. Suppose that there was data like Fig.5. It considers carrying out the modeling of the input-and-output relation from this data. The space is divided by the membership function on the coordinate axes of the two-dimension input. First, for each of the divided region, the data is plotted on a 2-dimensional plane (Fig.6 (a)). And, by applying IDS, the irradiation pattern is extracted (Fig.6 (b)). Considering that the brightness of light is a membership value, narrow path is extracted using the center-of-gravity method of de-fuzzy.

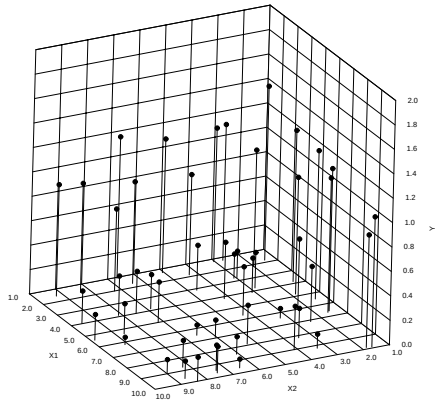
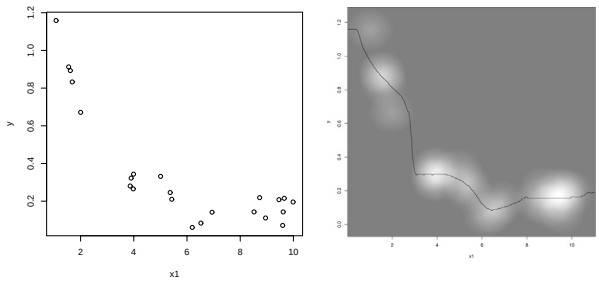


Fig.5 Data of 2in-1out



(a) plot (b) narrow path

Fig.6 Projection of data

On each of the coordinate axes, narrow path of each of the divided region is extracted. For each region, narrow path is expressed at a rule which used the membership function, like a formula (2-1).

$$\left.\begin{aligned} R_1 &: if\ x_2\ is\ A_{21},\ then\ y_{11}\ is\ U_{11} \\ R_2 &: if\ x_2\ is\ A_{22},\ then\ y_{12}\ is\ U_{12} \\ R_3 &: if\ x_1\ is\ A_{11},\ then\ y_{21}\ is\ U_{21} \\ R_4 &: if\ x_1\ is\ A_{12},\ then\ y_{22}\ is\ U_{22} \end{aligned}\right\} \quad (2\text{-}1)$$

$$y\ is\ \beta_{11}y_{11}\ or\ \beta_{12}y_{12}\ or\ \beta_{21}y_{21}\ or\ \beta_{22}y_{22} \quad (2\text{-}2)$$

The output y obtained by combining $y_{11}, y_{12}, y_{21}, y_{22}$ output of each rule by formula (2-2). β_{ij} is the weight when combining each narrow path, and is given according to the degree of influence in an output. Specifically, it determines as a quantity in inverse proportion to it from the size of a spread of narrow path. As mentioned above, a system-wide model is constituted. These process are shown in Fig. 7.

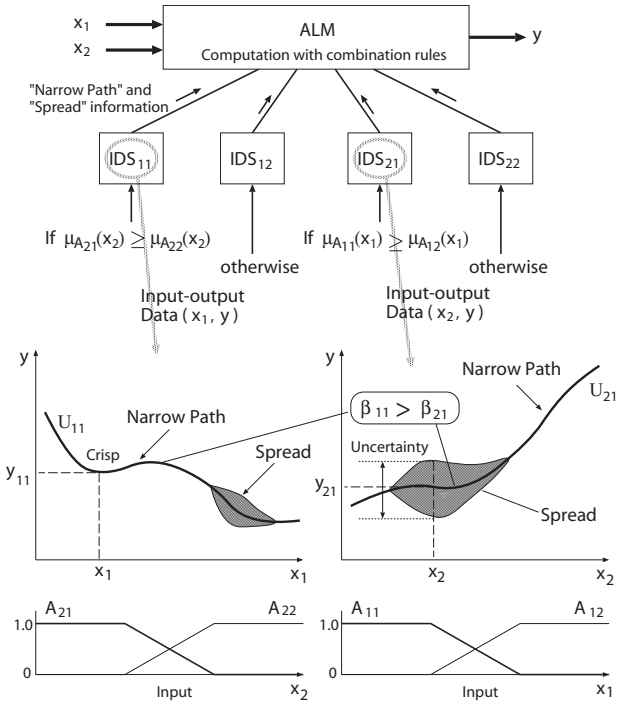


Fig. 7 Process of combining

C. Unsupervised Learning

When considering what is called study, there is a surely required thing. It is Evaluation. If what whose being good and being bad and this are not known, it does not understand in a thing that what should be learned. It is defined as study as optimizing so that this evaluation may be made into the maximum. When considering this evaluation in machine learning, in with a supervised data, evaluation is included in this data itself. However, it is necessary to hold the evaluation in the inside of one's system in the case where it have no supervised data. Therefore, ALM have an evaluation. Furthermore, in order to learn actively, ALM influences itself and generates the data to evaluate. An information gathering rule is used for the determination of this output, and it is determined probable based on the information on narrow path and its spread.

In order to realize the action with high evaluation, the controller must choose the action which was tried and proved successful in the past. But in order to find out the action with higher evaluation, it must choose the action which has not been tried before. In order to do action with high evaluation, the action must be decided using previously obtained knowledge, but, search for unknown states is necessary to obtain better action. For this purpose, search is done based on the presently best action and by changing it by probability.

The input-output system modeled by use of IDS is the form like distribution function of certainty value (Fig.8). By use of this spread, the input-output relation can be decided centering on the narrow path by probability. Thus, the output is decided based on the information of spread, and the result is evaluated, and modeling is repeated using the data with high evaluation.

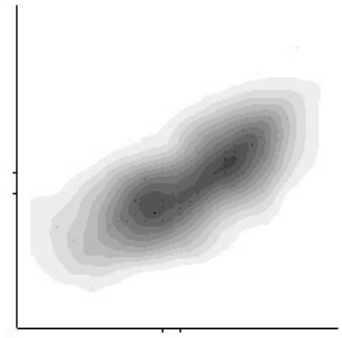


Fig. 8 SISO system

D. Supervised Learning

When there is supervised data, a rule can be gained by carrying out the modeling of the supervised data. However, the supervised data needs to be optimal or semi-optimal in that case. However, the optimal supervised data cannot be obtained in many cases in fact. Then, if supervised data is obtained, using it, supervised learning will be performed first and it will learn by unsupervised learning by making it into an initial state.

Therefore, when there is optimal teacher data, by carrying out the modeling of this, the optimal rule can be gained and the more optimal rule can be found out by carrying out unsupervised learning further. Moreover, even when only imperfect supervised data is obtained, by using the data as a key it can learn at high speed rather than it learns it from zero by searching the rule optimal.

III. Simulation

ALM performs the amplitude increase control simulation of double pendulum, and comparison with the case where the case where it learns from zero, and imperfect teacher data are learned as an initial value is performed.

A. Double Pendulum

Double pendulum called acrobot is a dynamics model similar to iron bar movement of human (Fig.9). Link 0 equivalent to the hand of holding an iron bar are a non-driving joint, and require only the dynamic friction torque. Control torque is given only to link 1 equivalent to the waist.

For this reason, theta 0 is dependent on the center-of-gravity position and posture form of a model. That is, if theta 1 is decided, the value of theta 0 cannot be decided arbitrarily.

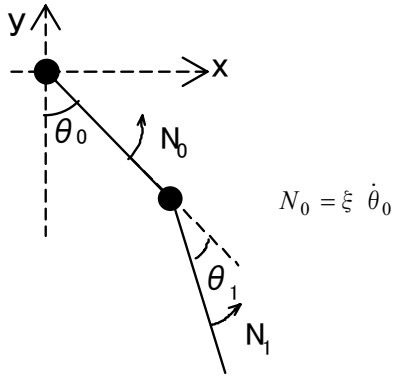


Fig. 9 Double pendulum
ξ is coefficient of viscous friction

Moreover, in which the link 1 has prepared drive restrictions in order to bring close to the model of iron bar movement more (-130 degree <= theta1 <= 130 degree) -- therefore, a link 1 cannot be rotated. Theta i is Relative angle of the link i-1 in each joint i and Link i. N is torque added to each link.

The kinetic energy is T (formula 3-1) and the potential energy by gravity is V (formula 3-2), and the kinetic equation is obtained from the Lagrangian equation (formula 3-3) to θ0...θ4 by the Lagrangian function L=T-V.

$$\mathrm{T} = \sum_{i=0}^{1} \left(\frac{1}{2} m_i (\dot{x}_i^2 + \dot{y}_i^2) + \frac{1}{2} I_i \dot{\theta}_i^2\right) \qquad (3\text{-}1)$$

$$V = \sum_{i=0}^{1} m_i g y_i \qquad (3\text{-}2)$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}_i}\right) - \frac{\partial L}{\partial \theta_i} = N_i \qquad (3\text{-}3)$$

Where x i, y i is the coordinates of center of gravity of the ith link and I i is the moment of inertia. The kinetic equation is dispersed by the four-dimensional RungeKutta method and computer simulation is done with Δt=1/1000 second.

B. Double Pendulum Control Simulation

The knowledge acquisition simulation system which gains the amplitude increase control rule using ALM is built by making the model of double pendulum control object.

Fig. 1 is applied, with the amount of operations (torque) which is the model of double pendulum and is received from a controller, the candidate for control will change the state (the angle of each link, angular velocity) of a model, and will return the state of the model to a controller.

At first, the controller outputs the amount of operations according to the information gathering rule, it collect The input-and-output data for control, it learns the control rule in the knowledge acquisition part based on the data.

When learning progressed, The amount of operations is outputted by considering the state of a model as an input by the control rule learned. The flow is as follows.

1 which in the study from zero outputs the amount of operations at random and collects input-and-output data. Supervised data is used when there is supervised data.
2 With an evaluation function, evaluation is given and the high operation result (input-and-output data) of evaluation is saved (it saves in a database by making into control knowledge what has the largest deflection angle).
3 The database of this control knowledge is changed into the form of a control rule, IDS is performed, and the modeling of the input-and-output relation is carried out.
4 Let this thing that carried out modeling be the control rule of a new control part.
5 According to a control rule, it opts for an output probable.
6 2-5 is repeated.

C. The Contents of an Experiment

Control object : double pendulum
Control purpose : amplitude increase
Control target : link 0 amounts to 180 degrees. (it makes perpendicular down into 0 degrees)
Mode1 : learning from zero.
Mode2 : learning using imperfect supervised data as an initial value. (rule which imperfect supervised data can reach to 100 degrees)

Mode1 and Mode2, an experiment is conducted 5 times, and the number of times of trial concerning learning is compared

D. Result

Fig. 10 is an example of a learning curve. it takes about 15000 times in .Mode1 (study from zero). It learned in Mode2 (study using imperfect teacher data) by about 6000 times and about 40% of number of times of Mode1. Even if it compares by the case of 180 degrees from 100 degrees, Mode1 takes about 10000 times and Mode2 is learning by about 6000 times and it is about 60% of number of times of Mode1.

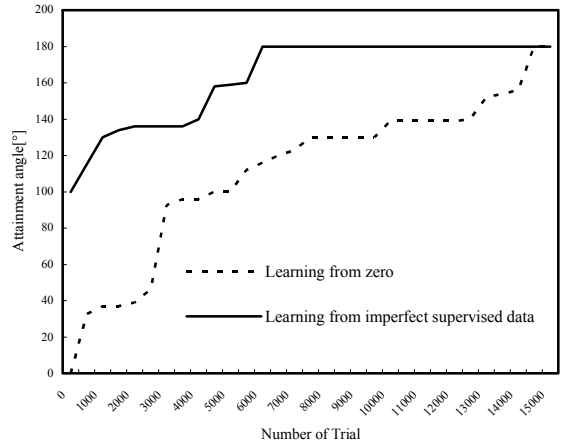


Fig. 10 Learning curve

However, which cannot be evaluated by one example since the learning by trial and error is random reference and learning time has change each time. Then we shows average value in Fig. 11.

Even if compared by average value, improvement in the learning speed by Mode2 using imperfect supervised data also can be seen. Mode2 takes about 59% of Mode1 in case of 0 degrees to 180 degrees. Mode2 learned by 70% of number of times of Mode1 in case of 100 degrees to 180 degrees.

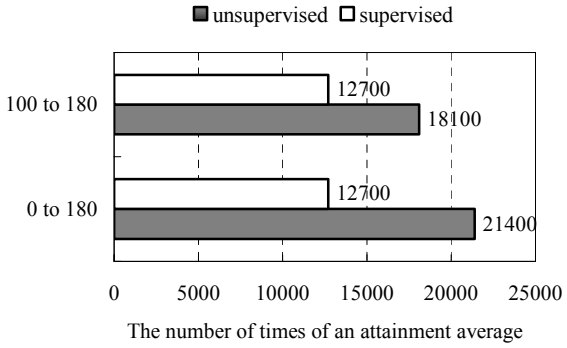


Fig. 11 average value

Since the state function which calculates the optimal value is not a convex function, when the supervised data given to an initial value is not near the global optimal value, it considered that it cannot desire improvement in a learning speed like this case. And, since the learning by trial and error is random reference, learning time has change each time.

Since the candidate for control was comparatively simple and the given imperfect supervised data was also comparatively given near the global optimal value this time, it is thought that improvement in study speed was obtained. In the case of for more complicated control, it is thought that there are many cases of the partial optimal value with the imperfect supervised data far from the global optimal value obtained, and improvement in the learning speed like this time cannot be desired in that case.

IV. Conclusion

In this paper, by proposing the active learning method for acquiring control knowledge actively by trial and error, letting the learning simulation of the action policy of double pendulum pass, and learning by trial and error based on imperfect supervised data showed it learned efficiently and learning speed improved rather than having learned from zero.

References

[1] G.Yuasa, S.B.Shouraki, N.Honda, Y.Sakurai: "Applying an active learning method to control problems", Asian Fuzzy System Symposium 2000(AFSS2000), tukuba japan, pp572-577, (2000)

[2] Y.Sakurai, N.Honda, J.Nishino: "Acquisition of Knowledge for Gymnastic Bar Action by Active Learning Method", Journal of Advanced Computational Intelligence Intelligent Informatics, Vol. 7, No.1, pp.10-18, (2003)